

Appl. No. 10/056,300
Reply to Office Action of January 25, 2006

CLAIMS

- 1) (Currently Amended) A computer-based method for extracting specific content from a web page [document] comprising content, comprising the steps of:
creating set of selection envelopes, wherein each selection envelope is associated with at least one selection command for locating a particular portion of said content within said web page [document], wherein each selection command is a function configured to locate the particular portion of said content to be enclosed by a corresponding selection envelope, wherein each successive selection command narrows said content to be enclosed by a corresponding [the] selection envelope corresponding to the selection command; and
applying each selection command and its corresponding selection envelope to said content until said specific content from said web page [document] is enclosed in a particular one of the selection envelopes.
- 2) (Previously Amended) The computer-based method of claim 1 wherein said selection envelope comprises:
a begin marker and an end marker, which respectively define the beginning and end of said selection envelope.
- 3) (Previously Amended) The computer-based method of claim 1 wherein said at least one selection envelope comprises:
a parent envelope and a child envelope.
- 4) (Previously Amended) The computer-based method of claim 3 wherein said child envelope is nested within said parent envelope.
- 5) (Previously Amended) The computer-based method of claim 3 wherein said child envelope partially overlaps said parent envelope.

Appl. No. 10/056,300
Reply to Office Action of January 25, 2006

6) (Previously Amended) The computer-based method of claim 3 wherein said child envelope is completely outside of said parent envelope.

7) (Currently Amended) The computer-based method of claim 1 wherein at least one of said plurality of selection commands comprises:
a command based on web page [document] structure.

8) (Previously Amended) The computer-based method of claim 1 wherein at least one of said plurality of selection commands comprises:
a command based on a character pattern.

9) (Currently Amended) The computer-based method of claim 1 wherein at least one of said plurality of selection commands comprises:
a combined command based on both web page [document] structure and a character pattern.

Appl. No. 10/056,300
Reply to Office Action of January 25, 2006

10) (Currently Amended) A computer-based method for extracting desired content from a content source web page, comprising the steps of:

parameterizing a plurality of selection commands to operate on said web page [source], wherein each selection command is a function configured to locate a [the] particular portion of content within said content source web page to be enclosed by a corresponding selection envelope, wherein each successive selection command narrows the particular portion of said content source web page to be enclosed by the selection envelope corresponding to the selection command;

defining a parent selection envelope corresponding to a first selection command for locating a first content within said content source web page;

using the first selection command associated with the parent selection envelope to select the first content from said content source web page;

determining whether said first content is said desired content; [[and]]

extracting said first content if said first content is said desired content;

defining a child selection envelope corresponding to a second selection command for locating second content within said content source web page if said first content is not said desired content;

using the second selection command associated with said child selection envelope to select the second content from said content source web page;

determining whether the second content is said desired content; and

extracting the second content if the second content is said desired content.

11) (Cancelled)

12) (Cancelled)

13) (Cancelled)

Appl. No. 10/056,300
Reply to Office Action of January 25, 2006

- 14) (Currently Amended) The computer-based method of claim 10 wherein said web page [source] comprises:
a section of a web page [document].
- 15) (Currently Amended) The computer-based method of claim 10 wherein said step of parameterizing said at least one selection command comprises:
determining whether said web page [source] is structured or unstructured, and selecting said at least one selection command based upon this determination.
- 16) (Previously Amended) The computer-based method of claim 15 wherein said at least one selection command comprises:
a structure based command selected from the group including:
select by name commands, select by location commands, select by sibling relationship commands, select by attribute commands and select by counter commands.
- 17) (Previously Amended) The computer-based method of claim 15 wherein said at least one selection command comprises:
a character bascd command selected from the group including:
select text contain commands and select text matching pattern commands.
- 18) (Previously Amended) The computer-based method of claim 15 wherein said at least one selection command comprises:
a combined structure and character based command.